

Super Resolution Imagery

From Multi-frame Sequences with Random Motion

2 March 1998

J. Schuler*, D. Scribner, M. Kruer
Naval Research Laboratory, Code 5636
Washington, D.C. 20375

ABSTRACT

A general method is described for achieving super resolution imagery from multiple frame image sequences that contain motion. The method assumes low-noise, focal plane array imagery recorded with uncontrolled image motion that can include some random jitter. Using this approach, moderate resolution, fast frame image sequences can be processed to achieve high-resolution image sequences displayed at conventional frame rates. The super resolution processing depends only on the imagery, requiring no externally controlled micro-dither or *a priori* information such as the sensor motion or range to the background. Typical sensor stabilization requirements are relaxed using this method, however, to achieve optimum performance there are some constraints on the motion. Specifically, the stabilization must still be good enough so that the resulting random dither is only a few pixels and must be statistically well behaved. Processing examples are given using previously recorded image sequences from a wide FOV MWIR staring array sensor on-board an aircraft.

1. Introduction

The concept of super resolution as a multi-frame image restoration problem has been known for a number of years.¹ A number of innovations to this restoration technique have been studied to improve the performance.² Nevertheless, most of the techniques pertained to cases in which the image motion was known or the motion that needed to be determined was a simple global shift or rotation. For many sensor systems, the resulting image motion can not be described as a simple global shift but can only be fully described by the optical flow function: a vector field quantity describing two shift dimensions for each pixel.

* On-site NRL contract employee affiliated with Unified Industries, Springfield VA.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 02-03-1998		2. REPORT TYPE Conference Proceedings		3. DATES COVERED (FROM - TO) xx-xx-1998 to xx-xx-1998	
4. TITLE AND SUBTITLE Super Resolution Imagery From Multi-frame Sequences with Random Motion Unclassified			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
			5d. PROJECT NUMBER		
6. AUTHOR(S) Schuler, J. ; Scribner, D. ; Kruer, M. ;			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Code 5636 Washington, DC20375			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS Director, CECOM RDEC Night Vision and electronic Sensors Directorate, Security Team 10221 Burbeck Road Ft. Belvoir, VA22060-5806			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT APUBLIC RELEASE					
13. SUPPLEMENTARY NOTES See Also ADM201041, 1998 IRIS Proceedings on CD-ROM.					
14. ABSTRACT A general method is described for achieving super resolution imagery from multiple frame image sequences that contain motion. The method assumes low-noise, focal plane array imagery recorded with uncontrolled image motion that can included some random jitter. Using this approach, moderate resolution, fast frame image sequences can be processed to achieve high-resolution image sequences displayed at conventional frame rates. The super resolution processing depends only on the imagery, requiring no externally controlled micro-dither or a priori information such as the sensor motion or range to the background. Typical sensor stabilization requirements are relaxed using this method, however, to achieve optimum performance there are some constraints on the motion. Specifically, the stabilization must still good enough so that the resulting random dither is only a few pixels and must be statistically well behaved. Processing examples are given using previously recorded image sequences from a wide FOV MWIR staring array sensor on-board an aircraft.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT Public Release	18. NUMBER OF PAGES 15	19. NAME OF RESPONSIBLE PERSON Fenster, Lynn lfenster@dtic.mil	
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified		19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number 703767-9007 DSN 427-9007	
				Standard Form 298 (Rev. 8-98) Prescribed by ANSI Std Z39.18	

An example of this case is a wide FOV sensor on-board an aircraft looking down at moderate slant angle. The upper section of the image may be looking toward the horizon where the relative motion is very small. The bottom of the image may be looking at the nearby terrain moving under the aircraft and moving relatively fast.

Assuming that the optical flow for each pixel can be accurately calculated and the sensor parameters are known (such as optical transfer function, FPA geometry, FPA detector response function, and FPA noise), then a super resolution image can be computed using multi-frame data. The method used to do the processing builds upon sampling theory and Wiener filtering. Later in this paper some examples are given of super resolution processing applied to airborne MWIR image sequences taken with wide FOV optics.

2. Super Resolution Using Controlled Micro-Dithering

The dimensions of a focal plane array does not necessarily limit the spatial resolution of a sensor. A sub-pixel dither scan of a scene will generate an image cube which can be used to reconstruct a single super resolution composite with spatial resolution above the Nyquist limit as defined by the detector center-to-center spacing on the FPA. Figure 1 shows a simple micro-dither scan pattern.

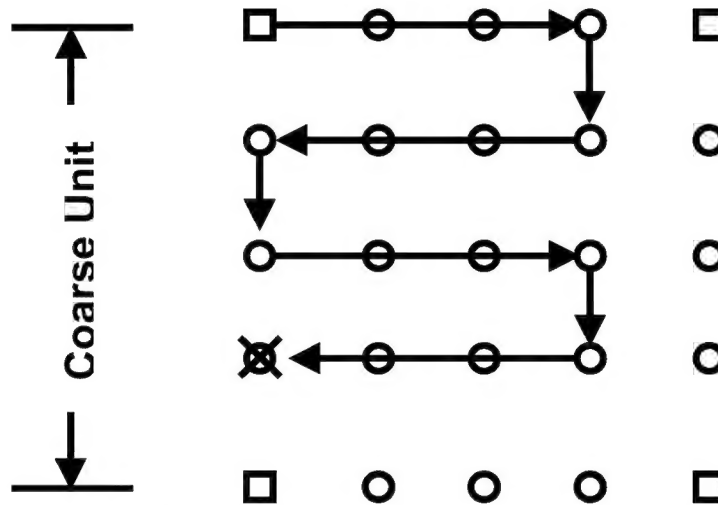


Figure 1: Typical controlled micro-dither scan pattern that achieves a 4 x 4 over-sampling.

In its simplest form, micro-dithering requires a perfectly stable sensor and accurate mechanical control of a scanning device. A number of mechanical devices are available that can generate high speed precise dithering as shown in Figure 1. A piezo-electric driver is an example of one such device. Using this technique on a moving platform with intrinsic jitter is not practical, as the micro-dither device alone no longer controls the motion of the image on the focal plane. The motion of the aircraft and the any associated vibration would predominate over that of the micro-dither scan pattern.

Nevertheless, it is instructive to examine the case of perfectly known, uniform micro-dither so that sampling issues of super resolution can be compared against interpolation. Increasing the number of pixels in a digital image involves changing the spatial sampling rate of the image on the focal plane. Controlled dithering increases the sampling rate by creating new sample points as shown in Figure 1. Alternatively, values at the new sample points could be estimated by interpolation of a single frame. However, such new samples are created using only the original information from the image. Interpolation is inherently band limited and adds no new information to the increased bandwidth afforded by a higher sampling rate are not utilized. This is shown graphically in Figure 2 below.

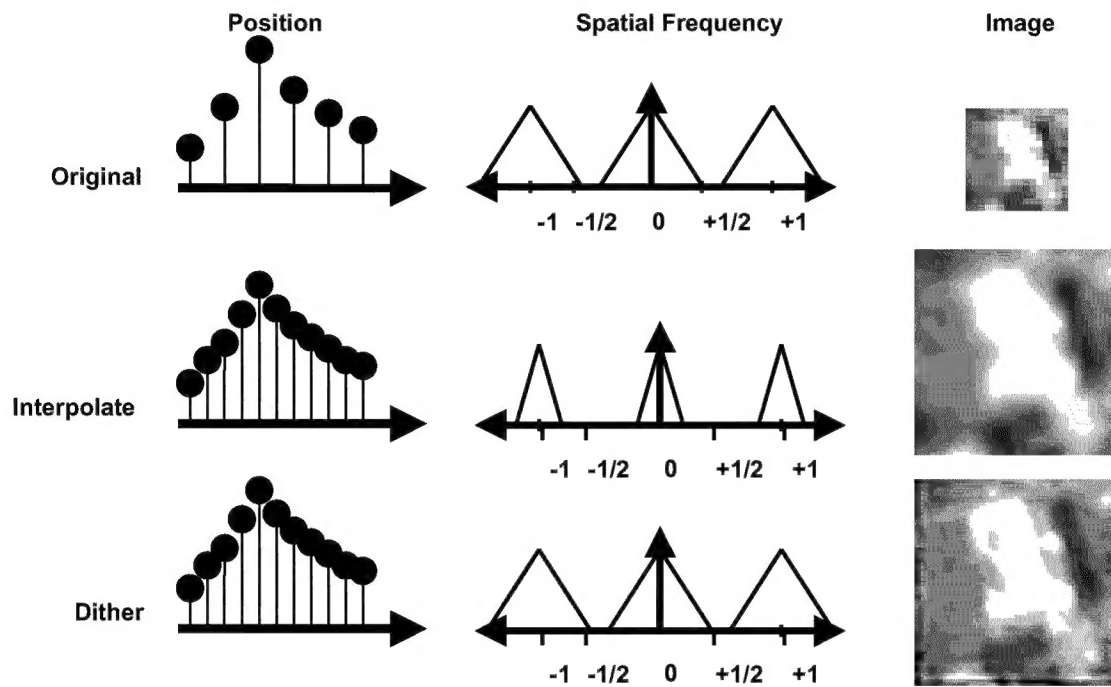


Figure 2: Comparison of interpolation and dithering methods to increase image size. Interpolation is band limited to the information of a single frame and results in a blurred image. Dithering increases sampling density through multiple exposures. Only through the combination of multiple frames can sharper features emerge that were not observable in any single image.

3. Super Resolution Using Random Motion

As described in the previous section, most super resolution methods rely on controlled dither scan patterns followed by multi-frame image reconstruction. A more utilitarian approach would be one that accepted small, random dither shifts between images in the multi-frame sequence. As before, the approach still requires a multi-frame sequence of shifted images, but the requirement for a precise, uniform micro-dither is relaxed. A simple description of the approach can be given as four steps. First, estimate the motion of the image in terms of the optical flow for each frame and each pixel in a multi frame sequence. Second, assemble the resulting multi-frame data into a single super resolution image on irregularly spaced sample locations. Third, interpolate the irregularly spaced samples onto a regularly spaced grid. Fourth, compensate for known blurring effects due to the optics and detector response function. Each step is discussed separately below.

3.1 Motion Estimation

The shift between two frames in an image sequence can be estimated by various methods. Generally, this motion will not be uniform across an image and requires computing the optical flow. Such a computation finds a shift for every pixel between two sequential frames. A demonstration of an optical flow calculation is given below in Figure 3. Our particular approach used phase correlation on a regional basis across the image with limited performance. This was improved by modeling an assumption that for well-behaved backgrounds and image motion, the optical flow will change monotonically across each frame pair. There exist in the general literature extensions of this technique to multi-scale or wavelet based implementations that promise increased computational robustness and efficiency.

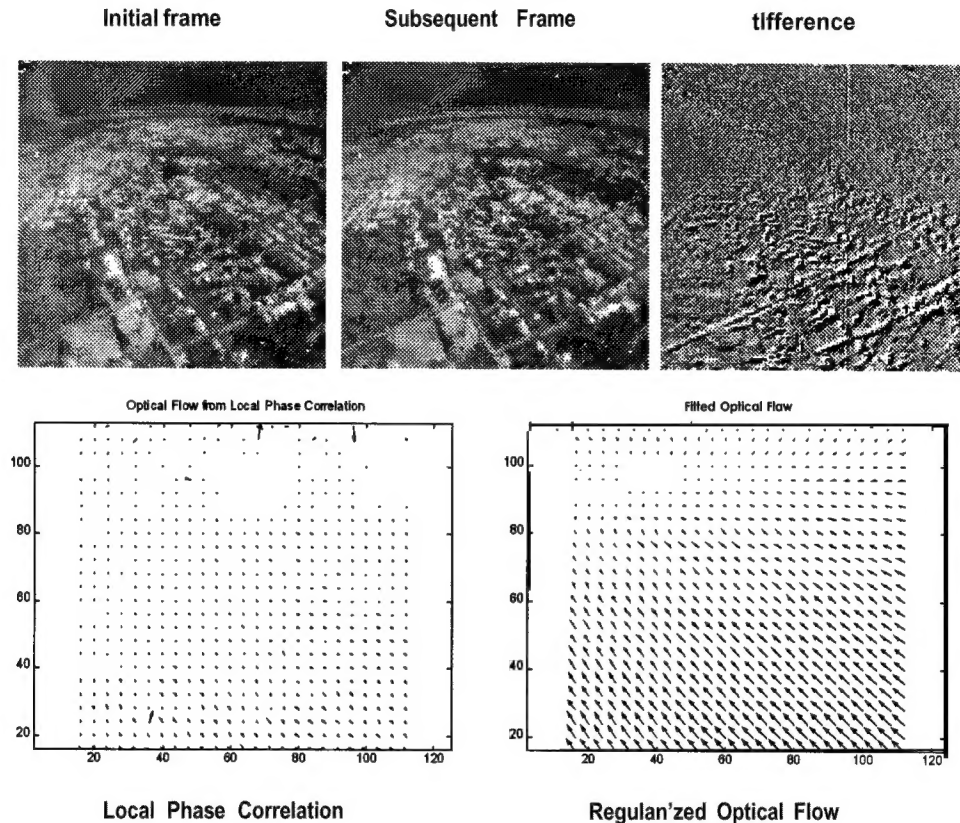


Figure 3: Sequential imagery from 50 Hz wide FOV 128x128 airborne imagery. The algebraic difference between frames demonstrates motion non-uniformity across the scene. Regional (32x32) shift estimators suffer from poor performance, but simple model fitting provides markedly improved motion estimation.

Note that in Figure 3 the motion, or optical flow is greatest at the bottom of the image and very small near the top of the image. In fact, it can be inferred from the optical flow plot that the aircraft is flying toward a point in the image defined by zero optical flow.

3.2 Assembling a Single Super Resolution Image from Multi-frame Data

With precise estimates of the optical flow for each frame of data, the original image sequence is combined into a single super resolution image by assembling pixels based on their resulting coordinates. Note that in doing so under random motion, the true pixel coordinates no longer lie on a rectilinear grid, yielding a non-uniform sampling of the optical image projected at the focal plane. A hypothetical example

of random dither points with respect to the original unit cell and the fine mesh super resolution grid is given below in Figure 4.

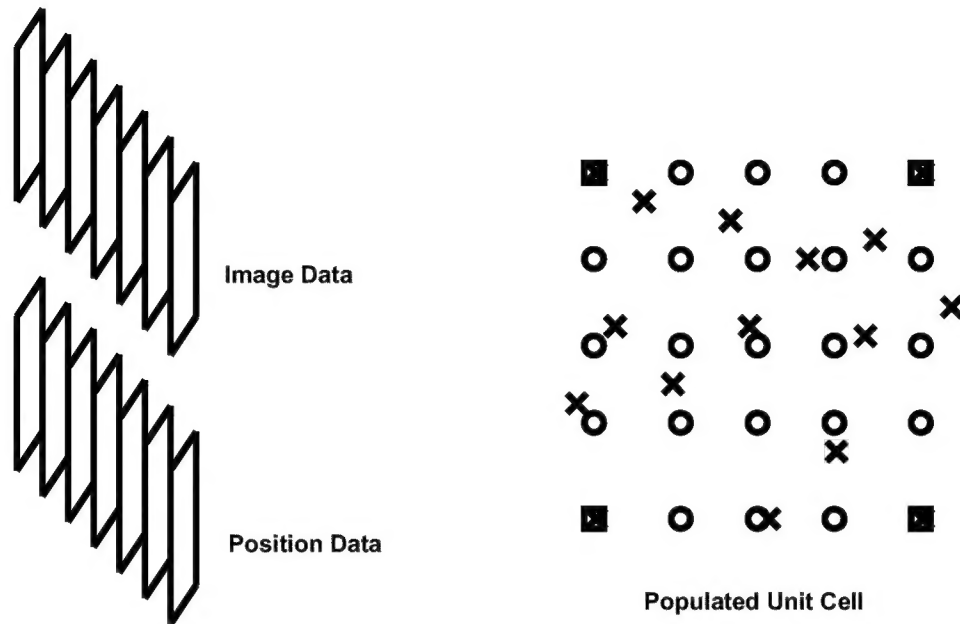


Figure 4: Graphical demonstration of assembling a 3-D spatial-temporal image deck into a composite given coordinate knowledge for every pixel in space and time. This knowledge can be successfully derived from motion analysis of the digital video, and requires no external control or cues.

Note that the distribution of points in Figure 4 characterize a relatively good distribution of sample points in space; That is for each fine mesh points on the regular grid (denoted by the open circles) there are several candidate dither points (denoted by the X's) in the immediate neighborhood. These nearby dither points can then be used to estimate intensities sampled at the regular lattice points. The actual distribution of dither points is not guaranteed to provide good cover over all parts of the image. For example, Figure 5 shows four different distribution from each corner of the same image sequence. Note that the two bottom corners are fairly well distributed. However the top left corner suffers from too little motion and top right corner suffers because the motion is restrict to be primarily in the horizontal direction.

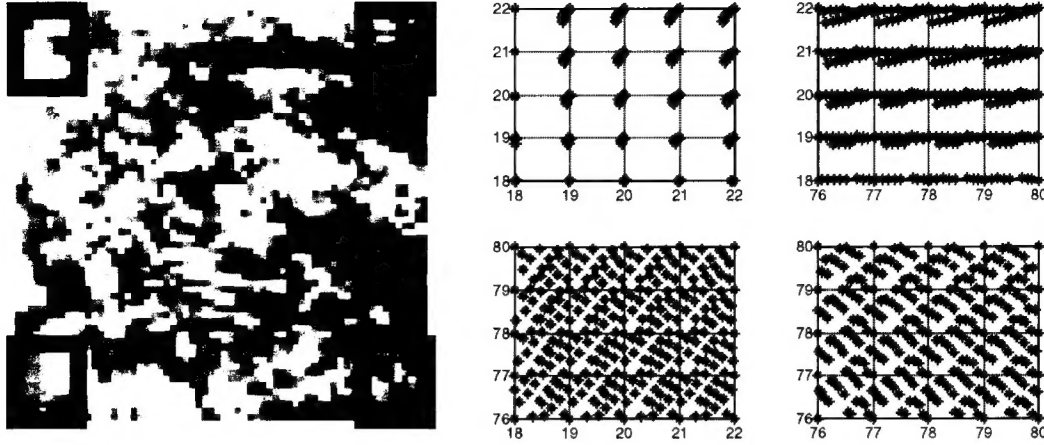


Figure 5: Pixel drifts for 16 frames of 50 Hz airborne wide FOV video. Note the pronounced regional dependence of drift motion during this episode. Pixels in the foreground drift across 3 coarse lattice sites, providing good coverage of fine lattice sites for super resolution. Other pixels drift minimally and offer little additional information than that of a single frame. Image quality of the super resolution composite tracks with the suitability of pixel drifts.

3.3 Regularizing the Super Resolution Image

Our approach is tolerant that the dither coordinates of the assembled composite do not lie on a rectilinear grid. This requires converting a non-uniformly sampled optical image to a regular sample array through interpolation. Our technique used a bilinear method based on the four nearest candidate dither intensities.

All interpolation techniques distort signal quality; such distortion degrades with increasing geometric distance between known data positions and estimated data positions. In the case of sufficiently high dithering density with small geometric interpolation distances, this step can be skipped when interpolation errors would be on the order of the sensor noise.

3.4 Recovering from known blurring effects

An image sampled by a focal plane array suffers from heavy blurring due spatial integration of each detector and the blur associated with resolution limited optics.⁴ The high spatial frequencies lost from blurring can be partially recovered by a restoration filter. In the derivation below, a simple model of optical and detector blurring is used to compute a de-convolution filter. We first model the optical image as a geometric projection of a real world scene onto a focal plane

$$s_g(\chi_1, \chi_2, t) = \left(\frac{\pi}{4} \frac{\tau_{op}}{F_{no}^2}\right) \Psi\left(\frac{\chi_1}{m}, \frac{\chi_2}{m}, t\right)$$

Next, we model the effect of OTF blur and spatial integration of the detector as linear operators acting on the geometric image to generate a pseudo image.

$$s_p(\chi_1, \chi_2, t) = h_{OTF}(\chi_1, \chi_2) ** h_{DET}(\chi_1, \chi_2, t) ** s_g(\chi_1, \chi_2, t)$$

$$S_p(F_1, F_2, t) = H_{OTF}(F_1, F_2) H_{DET}(F_1, F_2, t) S_g(F_1, F_2, t)$$

The effect of detector spatial integration can be modeled as convolution with a boxcar function.

$$s_p(\chi_1, \chi_2, t) = \frac{1}{a_1 a_2} \int_{\chi_1 - \frac{a_1}{2}}^{\chi_1 + \frac{a_1}{2}} \int_{\chi_2 - \frac{a_2}{2}}^{\chi_2 + \frac{a_2}{2}} s_g(\xi, \eta, t) d\xi d\eta = h_{DET}(\chi_1, \chi_2, t) ** s_g(\chi_1, \chi_2, t)$$

$$h_{DET}(\chi_1, \chi_2, t) = \text{rect}\left(\frac{\chi_1}{a_1}\right) \text{rect}\left(\frac{\chi_2}{a_2}\right)$$

$$H_{DET}(F_1, F_2, t) = a_1 a_2 \text{SINC}(F_1 a_1) \text{SINC}(F_2 a_2)$$

The optical system can be modeled with resolution limited spatial frequency response

$$H_{OTF}(F_1, F_2) = \text{rect}\left(\frac{F_1}{w_1}\right) \text{rect}\left(\frac{F_2}{w_2}\right)$$

$$h_{OTF}(x_1, x_2) = w_1 w_2 \text{SINC}(w_1 x_1) \text{SINC}(w_2 x_2)$$

Sampling the pseudo image by a regular lattice of detectors creates a discrete space response related to the continuous space response by

$$s_d[n_1, n_2, (t)] = s_p(n_1 b_1, n_2 b_2, t)$$

Both the continuous space and discrete space image can be characterized by their respective Fourier transforms

$$S_p(F_1, F_2, t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} s_p(\chi_1, \chi_2, t) e^{-i2\pi F_1 \chi_1} e^{-i2\pi F_2 \chi_2} d\chi_1 d\chi_2$$

$$S_d(f_1, f_2, t) = \sum_{n_1=-\infty}^{+\infty} \sum_{n_2=-\infty}^{+\infty} s_d[n_1, n_2, (t)] e^{-i2\pi f_1 n_1} e^{-i2\pi f_2 n_2}$$

The continuous space Fourier transform and discrete space Fourier series are related by

$$S_d(f_1, f_2, t) = \frac{1}{b_1} \frac{1}{b_2} \sum_k \sum_l S_p\left(\frac{(f_1-k)}{b_1}, \frac{(f_2-l)}{b_2}, t\right) \quad f_1 = \frac{F_1}{F_{\text{sample}}} = \frac{F_1}{\frac{1}{b_1}} = b_1 F_1$$

For an array of detectors with size a and pitch b , we present graphical relationship between space and spatial frequency domains

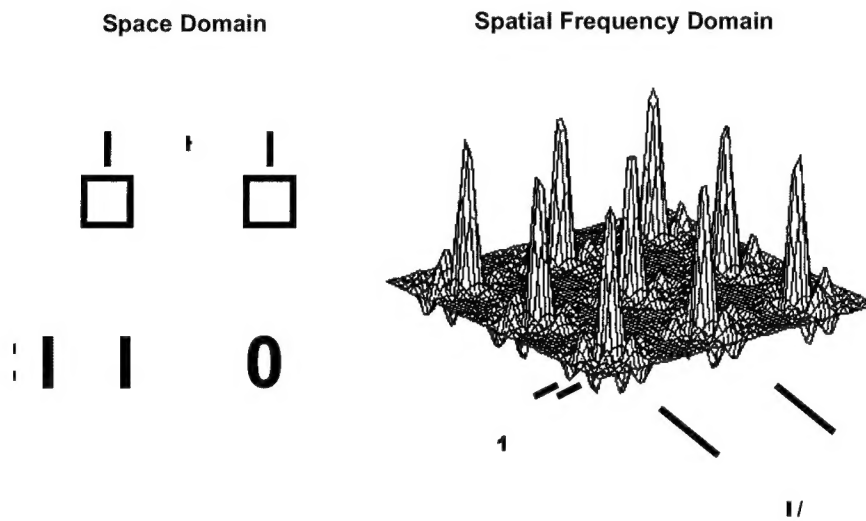
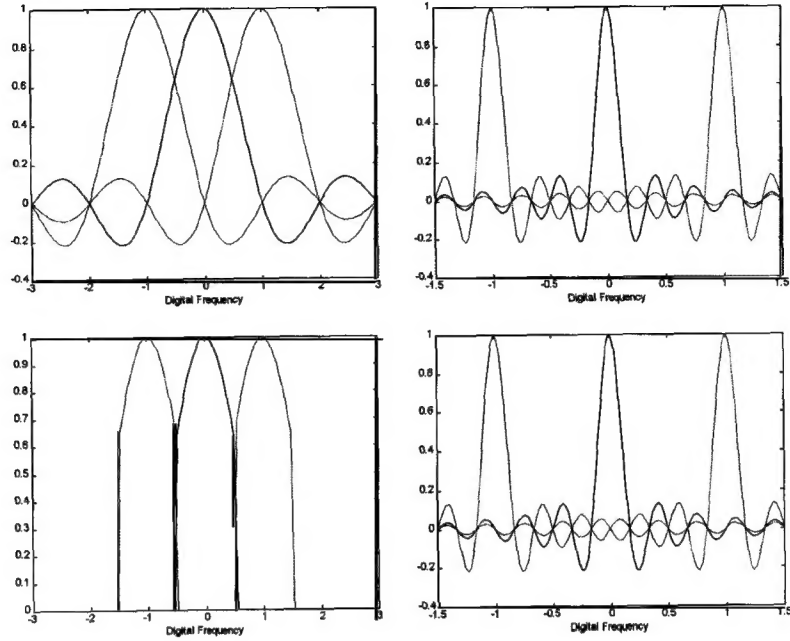


Figure 6: Relationship between the spatial response of an ideal integrating FPA and the corresponding modulation in spatial frequency space

Ignoring optical blur, a 100% fill factor focal plane array corrupts the sampled image with heavy alias distortion as shown in Figure 7 (upper left). By applying diffraction-limited optics with at least 2 pixels per blur circle, alias errors are eliminated at the cost of reduced spatial resolution (lower left). Increasing the sampling rate through dithering reduces alias error by further separating the discrete spatial frequency images of the pixel blur function (upper right). Applying a diffraction limiting optic of smaller blur circle can again eliminate alias distortion (lower right).



**Figure 7: (Upper Left) Spatial frequency of a 100% fill factor FPA
 (Lower Left) 100% fill factor FPA with Nyquist blurring optic
 (Upper Right) 3-fold dithered 100% fill factor FPA
 (Lower Right) IT-fold dithered 100% fill factor FPA with Nyquist blurring optic**

Dithered imagery still suffers from heavy blurring due to detector spatial integration. Such linear effects can be optimally restored with a Wiener filter given the spectra of image clutter and sensor noise.

$$G_{Wiener}(f_1, f_2) = \frac{H_{blur}^*(f_1, f_2) P_{signal}(f_1, f_2)}{|H_{blur}(f_1, f_2)|^2 P_{signal}(f_1, f_2) + P_{noise}(f_1, f_2)}$$

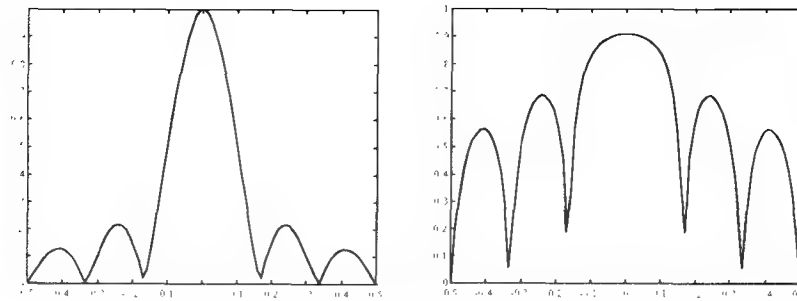


Figure 8: Ideal distortion modeling of a 100% fill factor dithered FPA with Nyquist optic. On the right is the optimal least squares (Wiener) restoration filter given a 10:1 SNR. Dithering allows recovery of pixel blurring because alias distortion is pronouncedly reduced

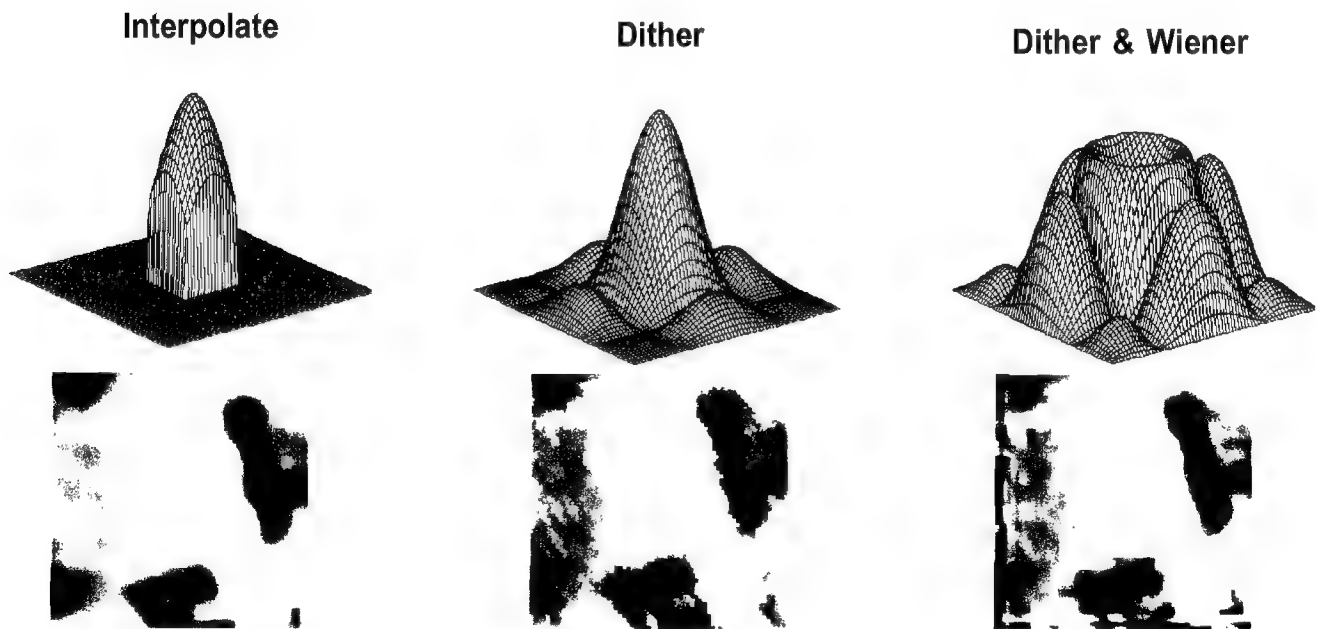


Figure 9: Modeling of the spatial frequency response and resultant imagery, demonstrating interpolation, dithering, and dithering followed with Wiener filter restoration.

the sensor is sufficiently dithered that the alias distortion is on the same order as the sensor noise. Depending on OTF blur and clutter-to-noise spectra, meaningful dithering is roughly limited to a resolution of the focal plane array dimensions multiplied by the signal to noise ratio.

4. Results

Our procedure was implemented on NRL airborne mid wave infrared data taken August 1995. The original digital video was wide FOV, 128x128 pixels clocked at 50 frames/second. Our demonstration integrated 16 frames of data using only the ambient aircraft motion to dither the sensor. The performance of our super resolution algorithms heavily depends on ambient motion, which varies dramatically across the field of view of the sensor. Enhancements and limitations are clearly observable in the resultant imagery presented in figures 10-13, and closely track with local scene drifts. Every figure shows matching pairs of 64x64 cropped and zoomed imagery that demonstrate the performance of interpolation and super resolution methods compared with the original low resolution scene.

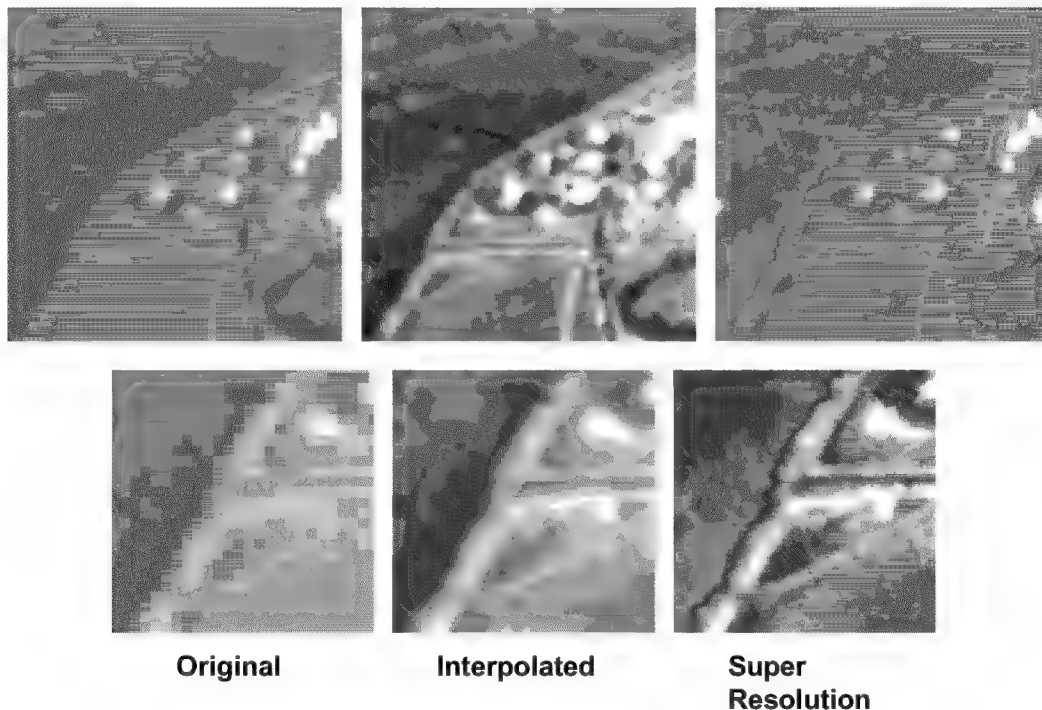


Figure 10: Comparison of 4-fold interpolation and 4X super resolution of a 64x64 image

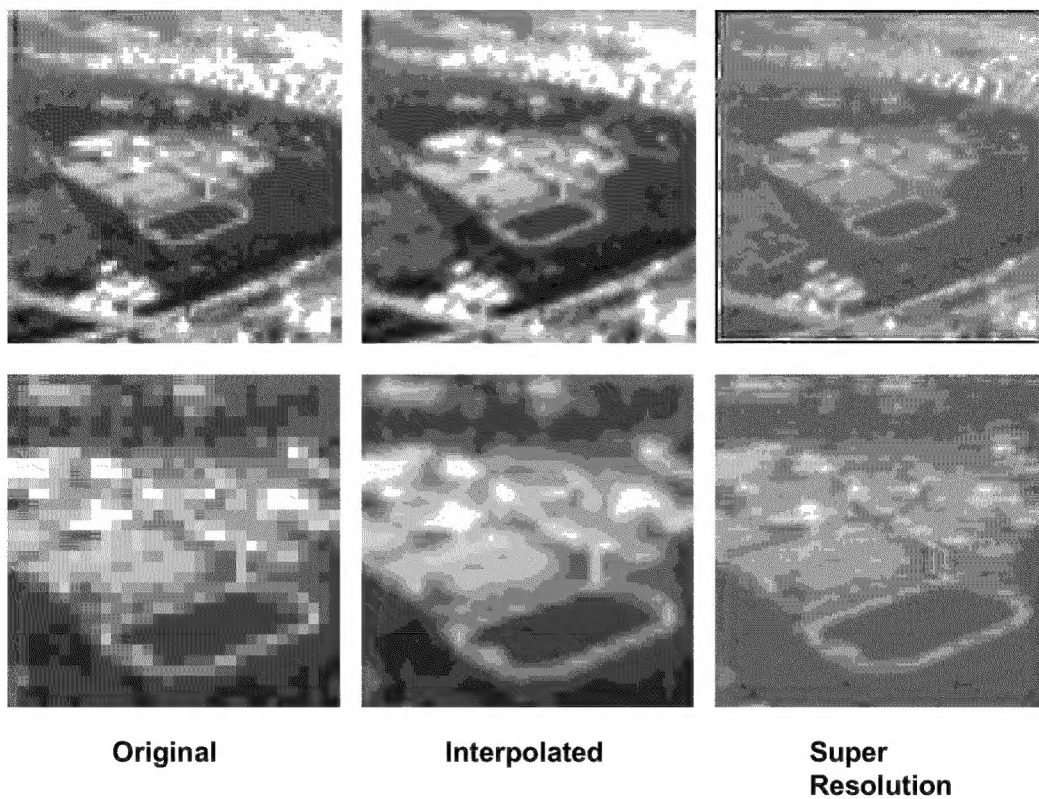


Figure 11: Comparison of 4-fold interpolation and 4X super resolution of a 64x64 image

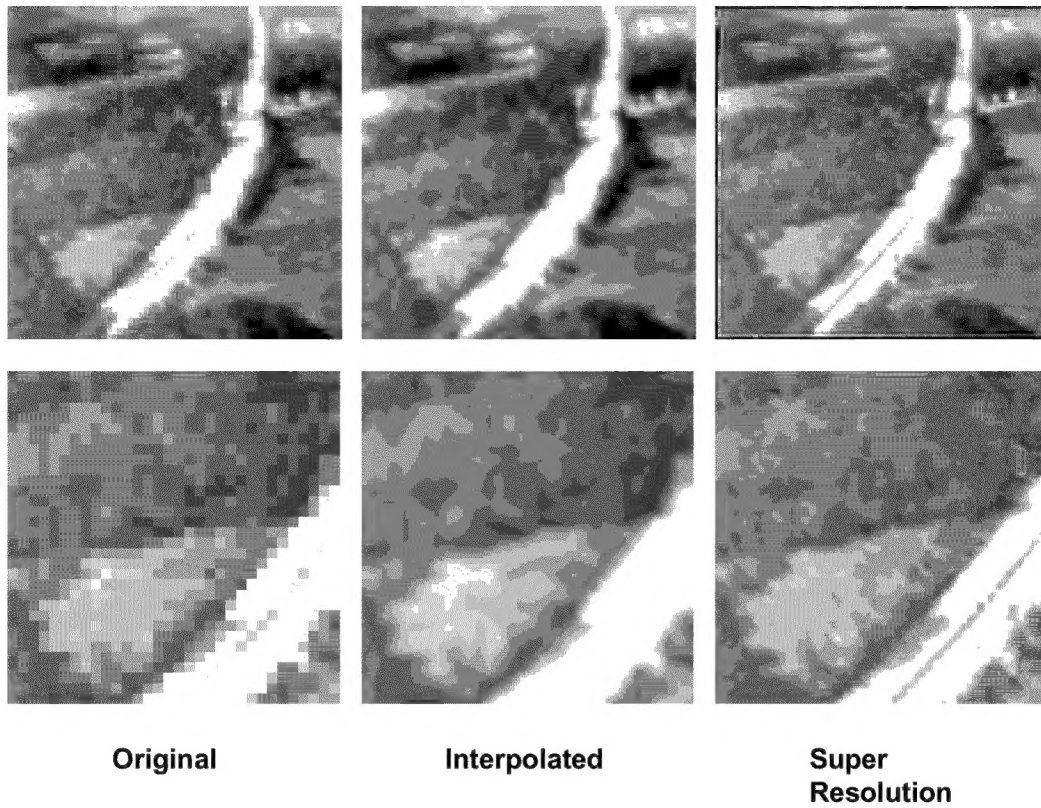


Figure 12: Comparison of 4-fold interpolation and 4X super resolution of a 64x64 image

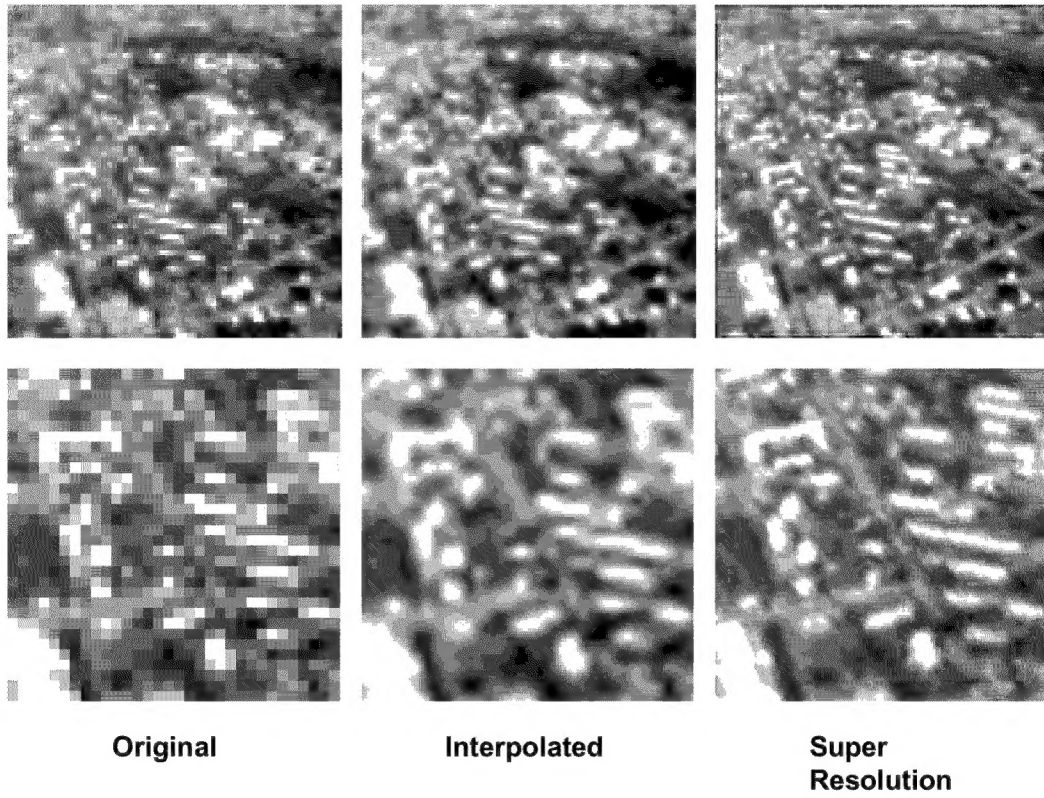


Figure 13: Comparison of 4-fold interpolation and 4X super resolution of a 64x64 image

5. Summary

A general technique is presented for generating higher resolution imagery from lower resolution image sequences containing uncontrolled scene motion. This technique estimates scene motion from a temporal optical flow calculation. Knowledge of pixel motion and intensities allows for the assimilation of a composite image of higher resolution than any single frame. Minor modifications to the data set account for the non-uniform spatial sampling intrinsic with uncontrolled motion, the performance of which depends on the stochastic distribution of the motion. A final restoration filter, dependent on sensor parameters and noise performance, is applied to the data set to provide an optimal recovery of known distortions intrinsic to super resolution. This process is demonstrated on 50 Hz midwave infrared imagery from a hard-mounted staring sensor flown around 3,000 feet altitude. The wide range of situational motion across the scene variably impacted final performance of the algorithm.

¹ T. S. Huang and R. Y. , “Multiple frame image restoration and registration”, in *Advances in Computer Vision and Image Processing*, vol. 1, JAI Press, 1984.

² M. Elad and A. Feuer, “Restoration of a single superresolution image from several blurred, Noisy, and undersampled measured images”, *IEEE Trans. on Image Processing*, vol. 6, p. 1647, 1997.

³ A. Schaum and M. McHugh, “Analytic methods of image registration: displacement estimation and resampling”, NRL Report 9298

⁴ D. A. Scribner, M. S. Longmire, and M. R. Kruer, “Analytic modeling of staring infrared systems with multidimensional matched filters”, *SPIE Vol. 890 Infrared Systems and Components II* (1988)

⁵ A. K. Jain, “Fundamentals of Digital Image Processing”, Prentice Hall, 1989